

Europäisches Patentamt

European Patent Office

Office européen des brevets

(1) Publication number:

0 079 739

A2

12

# EUROPEAN PATENT APPLICATION

(21) Application number: 82305926.6

(22) Date of filing: 08.11.82

(51) Int. Cl.<sup>3</sup>: C 12 N 15/00

C 12 N 1/00, C 12 P 21/02 C 07 H 21/04, C 07 C 103/52 //C12R1/19, C12R1/865

30 Priority: 12.11.81 US 320632

(43) Date of publication of application: 25.05.83 Bulletin 83/21

(84) Designated Contracting States: BE CH DE FR GB IT LI NL SE (1) Applicant: THE UPJOHN COMPANY 301 Henriette Street Kolamazoo, Michigan 49001(US)

(7) Inventor: Dugaiczyt, Achilles c/o The Upjohn Company 301 Henrietta Street Kalamazoo Michigan 49001(US)

(4) Representative: Porry, Robert Edward ot al, GILL JENNINGS & EVERY 53-64 Chancery Lane London WC2A 1HN(GB)

(54) Albumin-based nucleotides, their replication and use, and plasmids for use therein.

(f) The DNA sequence coding for human serum albumin has been isolated and inserted as two fragments into two novel plasmids which can be replicated in *E. coli*. These novel fragments can be joined to provide a unitary DNA sequence which then can be cloned into a suitable host, e.g. *E. coli*, for the expression of human serum albumin (which is used extensively in medical practice in treating shock conditions).

738 P

6/0 0

<u>@</u>

UPJOHN 1 GJE 70/2056/02

ALBUMIN-BASED NUCLEOTIDES, THEIR REPLICATION AND USE, AND PLASMIDS FOR USE THEREIN

This invention relates to nucleotides related to human serum albumin (HSA), their replication and use, and plasmids (and host substances) for use therein.

The gene for serum albumin is regulated in

5 development. On the other hand, serum albumin is synthesised in mammals by the adult liver, and its plateau in adulthood. The embryonic liver and yolk sac, on the other hand, produce predominantly a-fetoprotein, but the synthesis decreases drastically after birth. Recently,

10 Law et al determined the complete sequence of mouse  $\alpha$ -fetoprotein mRNA, Nature 291 (1981) 201-205. The structure revealed extensive homology to mammalian serum albumin, indicating that the two proteins are encoded in the same gene family. Similar conclusions have been 15 reached from studies on the  $\alpha$ -fetoprotein genes of the

rat and the mouse; see Jagodzinski et al, Proc. Natl. Acad. Sci. USA, 78 (1981) 3521-3525, and Gorin et al, J. Biol. Chem. 256 (1981) 1954-1959.

The complete nucleotide sequence of human serum

20 mRNA has been determined from recombinant cDNA clones and
from a primer-extended cDNA synthesis on the mRNA
template. The sequence comprises 2,078 nucleotides,
starting upstream of a potential ribosome binding site
in the 5'-untranslated region. It contains all the

- 25 translated codons and extends into the poly(A) at the 3'-terminus. Part of the translated sequence codes for a hydrophobic prepeptide met-lys-trp-val-thr-phe-ile-ser-leu-leu-phe-leu-phe-ser-ser-ala-tyr-ser, followed by a basic propeptide arg-gly-val-phe-arg-arg. These signal
- 30 peptides are absent from mature serum albumin and, so far, have not been identified in their nascent state in humans. A remaining 1,755 nucleotides of the translated mRNA sequence code for 585 amino acids which are in agreement, with few exceptions, with the published amino
- 35 acid data for human serum albumin. The mRNA sequence verifies and refines the repeating homology in the tripledomain structure of the serum albumin molecule.

### DETAILED DESCRIPTION OF THE INVENTION

Human serum albumin cDNA is cloned into the PstI site of plasmid pBR322 by the oligo(dG)-oligo(dC) tailing technique. Plasmid DNA was isolated from 97 positive colonies which hybridized to the enriched albumin cDNA probe, and the recombinant plasmid pHA36 was found to contain the largest insert of an albumin cDNA sequence. Its restriction endonuclease map is shown in the drawing, together with a restriction map of the primer-extended plasmid clone pHA206. The latter was obtained in a second transformation experiment after initiating the cDNA synthesis from an internal primer. This primer was a 91 base pairs long DNA fragment, MspI(152)-TaqI(182/3), isolated from pHA36. The two plasmids, pHA36 and pHA206, share 0.15 kb of homologous DNA. Together, they encode the entire sequence for human serum albumin, starting with the CTT codon for leu -10 of the prepeptide and extending into the 3'-untranslated region of poly(A).

Sequence of the Albumin cDNA. The sequence was determined for the most part on both DNA strands to ensure accuracy. All of the restriction sites used to end-label DNA fragments were sequenced across by labeling a neighboring restriction site. The entire nucleotide sequence of the serum albumin mRNA, as determined from the cloned DNA in pHA36, pHA206, and from the primer-extended cDNA at the 5'-terminus of the message, is shown in the following Table 1. The inferred amino acid sequence is also indicated. The mRNA length is 2,078 nucleo-25 tides, of which 38 represent the 5'-untranslated region, 54 identify a prepeptide of 18 amino acids, 18 identify a propeptide of 6 amino acids, 1,755 code for the known 585 amino acids of serum albumin, 189 make up the 3'-untranslated region and 24 are the poly(A) sequence. Nucleotides 5 to 15 (-34 to -24) in the 5'-untranslated region (Table  $^{30}$  1) are complementary to a 3'-terminal region of eukaryotic 18S RNA [Azad, A.A. and Deacon, N.J. (1980) Nucl. Acids Res. 8, 4365-4376] and thus could represent a ribosome binding site:

(5')...T 
$$T^{C}$$
T C T T C T G T......albumin mRNA (3')...G A G G A A G G C G U C C  $m_{2}^{C}$ A  $m_{2}^{C}$ A......185 RNA

35

The translated portion of the mRNA sequence codes for the signal peptide and the main body of the albumin polypeptide chain. The

signal peptide is composed of a hydrophobic prepeptide of 18 amino acids and a basic propeptide of 6 amino acids (Table 1). Since prepeptides are removed from nascent secretory proteins (like albumin) in the endoplasmic reticulum, they are seen only in vitro in heterologous translation systems. As yet, they have not been found within cells [Judah, J.D. and Quinn, P.S. (1977) FEBS 11th Mtg., Copenhagen 50, 21-29; and Strauss, A.W., Donohue, A.M., Bennett, C.D., Rodkey, J.A. and Alberts, A.W. (1977) Proc. Natl. Acad. Sci. USA 74, 1358-1362]. This is the first report of the presence and the sequence of a prepeptide for human serum albumin. As it is with other secretory proteins, the conversion of proalbumin to albumin takes place in the Golgi vesicles, and the enzyme responsible for this cleavage is probably cathepsin B [Judah, J.D. and Quinn, P.S. (1978) Nature 271, 384-385]. This is also a first report on the sequence of the propeptide for normal human serum albumin.

At the 3'-end of the message, the putative polyadenylation signal sequence, AATAAA, is located 164 nucleotides downstream from the amino acid termination codon TAA and 16 nucleotides upstream from the beginning of the poly(A) sequence. Another characteristic sequence located near the polyadenylation site has been identified by Benoist, et al. [Benoist, C., O'Hare, K., Breathnach, R. and Chambon, P. (1980) Nucl. Acids Res. 8, 127-142]; the concensus sequence from several mRNAs was concluded as TTTTCACTGC. A similar sequence, TTTTCTCTGT, is located 19 nucleotides upstream from the AATAAA hexanucleotide in the human albumin mRNA (Table 1).

# TABLE 1

	(30)	(170)	(560)	(350)	(044)	(330)	(029)	(710)	(300)
5	aer AGC	20 1ys AAA	50 ala GCA	80 leu CTT	110 Pro CCA	140 try TAT	170 aln GAA	200 cys TCT	230 glu GAA
J	phe TTT	phe TTC	phe III	th. Act	asn AAC	130 asn glu qlu thr phe leu lys lys tyr leu try AAT GAA GAG ACA TTT TTG AAA AAA TAC TTA TAT	169 cys TGC		ala GCA
	leu CTC	asn	alu CAA	ala GCA	ase GAC	tyr TAC	168 cys TGT	leu CTC	
	phe TTT	95 68	thr	val	asp CAT	173 AAA	3 A	arg AGA	alu GAG
	leu CTT	91c	val GTA	thr	1 ya	lys AAA	thr ACA	<b>al</b> n CAG	ala CCT
10	-10 ser leu leu TCC CTT CTT	55 55	e A	25 233 10C	h1s CAC	leu TTG	phe TTT	1ys AAA	1 ys
	ser TCC	leu TTG	asn AAT	leu TTA	q In	phe TTT	ala GCT	213 CCC	5 T
	11e ATT	asp		1ys AAA	leu TTG	thr	e SCT		gln arg phe CAG AGA TTT
	phe TTT	1ya AAA	40 val lys leu val GTA AAA TTA GTG	ese CAC	phe	a) a	tyr lys Tat aaa	ser	gln arg CAG AGA
	p r o trp val tlu phe TGC GTA ACC TTT	phe TTT	1ys AAA	<b>4</b> 8	101 933 1GC	glu GAA	tyr Tat	ala CCT	
15	p r val CTA	10 819 000	40 val GTA	70 phe TTT	중 축 옷	130 asn AAT	160 arg AGG	190 1 ys	220 leu ser CTG AGC
	5 5	hls CAT	n is	leu CTT	asn AAT	asb GAC	ala lys CCT AAA	alu aly CAA GGG	leu CTG
	1ys AAG	ala GCT	esp CAT	thr	<b>P S</b>	h13 CAT	ala GCT	5 ₹	ဦ ဗိ
•	-18 Met CCTTTTCTCTTCTGTCAACCCCACAGCCCTTTGGCACA ATG	-1 1 arg asp ala his lys ser glu val CGA GAT GCA CAC AAG AGT GAG GTT	phe qlu TTT GAA	leu his CTT CAT	pro aly arq CCT GGG AGA	ala phe GCT TTF	phe phe TTC TTT	asp GAT	210 ala phe lya ala trp ala val ala arg GCT ITC AAA GCA TGG GCA GTA GCT CGC
20	CACA	g g	phe TTT	S E	Pro			620	val
	1160	lys ser AAG AGT	<b>6</b> 600	ser ICA	gin glu CAA GAA	124 oys thr IGC ACT	glu leu leu GAA CTC CTT	glu leu GAA CTT	trp ala TCC CCA
	CCCT	lys	34 938 1CT	1ys AAA	cAA CAA		leu CTC		15 E
	ACAG	ala his GCA CAC	gln gln CAG CAG	62 cys asp IGT CAC	1ys AAA	met ATG	glu GAA	asp GAT	lys sla AAA GCA
	ວັງວວ	ala GCA			<b>618</b> GCA	val	5 5 5 5 5 5 5 5 5 5 5 5 7 5 7 5 7 5 7 5	leu CTC	1ya
25	TCAA	asp GAT	30 tyr leu TAT CTT	60 glu asn GAA AAT	91 Cys TGT	asp	ele SCC	1ys	phe TTC
	TCTG	- ago	30 tyr TAT		8 6 7 7 7 7 7	120 val GTT	150 tyr TAT	8 5 2	
	CTCT	p r o y val phe arg o	gln CAG	ala GCT	asp CAC	glu GAG	phe TTT	177 ala oys leu leu GCC TGC CTG TTG	glu arg GAA AGA
	1111	r o phe	ala CCT	2	ala ccT	Pro CCA	tyr TAC	leu CTG	glu GAA
	ន	ral CTG	phe TTT	asp glu GAT GAG	met ATG	val arg GTG AGA	910 100	177 993 100	91y 86.
30		4. 19.	8 8 8 8 8	asp CAT	gly glu met GGT GAA ATG		his CAI	200 .	af E
		arg AGC	11e ATT	val ala GTT GCT	5.52	1 TG	arg AGA	ala GCT	leu gin lys CTC CAA AAA
			Jeu TTG	val CTT	tyr TAT	arg CGA	Ag ar	1ys	g al
		ala tyr GCT TAT	val crc	53 cys	glu thr tyr GAA ACC TAT	leu pro CTC CCC	ile ala arg arg his pro ATT CCC ACA ACA CAT CCT	asp CAT	ser leu gin lys phe AGT CTC CAA AAA TTT
35			21 ala leu val leu ile ala phe GCC TTG GTG TTG ATT GCC TTT	51 53 lys thr cys AAA ACA IGI		leu CTC	11e ATT	ala CCT	ser ACT
		ser ICG	21 ala GCC	2 2 8 8 8 8 8	81 arg CCT	asn AAC	glu GAA	171 ala CCT	201 ala GCC

	(890)	(980)	(1070)	(1160)	(1250)	(1340)	(1430)	(1520)	(1610)	(1700)
	260 Jeu CTT	290 11e ATT	320 ala Gri	350 ala GCC	380 Jeu CCT	410 819 CGT	\$\$0 h1s CAT	&70 ser AGT	500 1 ys AAA	530 val CTT
5	asp GAC	289 cys TGC	tyr TAT	leu CTT	pro CCT	val CTT	1ys AAA	val GTA	0 D	leu CTT
	arg ala Acc occ	ser his ICC CAC	asn AAC	ara AGA	phe lys TTT AAA	ala leu leu val CCC CTG TTA GFF	438 cys TGT	er CCA	val GTT	ala CCA
	arq		1ys AAA	leu CTG	phe TTT	Jeu CTG	437 cys TGT	thr	tyr TAC	thr
	asp GAC	glu lys CAA AAA	316 cys 10°	ser väl val leu leu leu arg leu TCT GTC GTG CTG CTG AGA CTT	g to	၁၂၁ ၁၂၁	lys AAA	glu lys GAC AAA	asp glu thr tyr val GAT GAA ACA TAC GTT	514 cys thr leu ser qlu lys qlu arq qln lle lys lys qln thr TCC ACA CTT TCT GAG AAG GAG AGA CAA ATC AAG AAA CAA ACT
10	asp GAT	glu GAS	asp val	leu CTG	asp CAT	qln əsn CAG AAT	ser AGC		ole GAA	erg gin lie lys lys AGA CAA ATC AAG AAA
	ala GCT	leu leu CTG TTG	asp CAT	val GTG	val phe GTG TTC	phe qln TTC CAG	4 55 55	leu his TTG CAT	asp CAT	1ys AAC
	253 cys TGT	leu CTG	1ys AAG	vál CTC	val GTG	phe TTC	val		val	11e ATC
	leu glu CTT GAA	lys pro AAA CCT	glu ser GAA AGT	ser TCT	ala iys val GCC AAA GTG	lys AAA	430 ser arg asn leu qly lya val qly TCA AGA AAC CTA GCA AAA GTG GAC	461 cys val IGT CTC	490 ala leu glu val GCT CTG GAA GTC	ag C&
•	leu CTT	1ys AAA		ty TAC		tyr	<del>2</del> 8	461 cys TGT	leu CTG	ara AGA
15	250 1eu CTG	280 91u CAA	310 phe val TTT CTT	340 88p GAT	370 171 171	400 41u GAG	430 leu CTA	460 gin leu CAG TTA	490 ala CCT	520 ser qlu lys qlu TCT GAG AAG GAG
	asp CAT	279 cys 1GT	phe TTT	pro CCI	369 alu cys CAA TGC	4 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8	asn AAC		ser ICA	1ys AAG
	245 246 glu cys cys his gly Gaa toc toc cat gca	278 0ys 1CC	asp CAT	arg his AGG CAT		leu CTT	arg AGA	val leu asn GTC CTG AAC	phe TTT	a Ju GAG
	ris Cat	lys leu lys glu AAA CTG AAG GAA	ala GCT	arg AGC	h is CAT	aln CAG	se <b>r</b> ICA	Jeu CTG	pro cys CCA TGC	ser TCT
20	245 246 cys cys TCC TCC	leu lys CTG AAG	ala GCT	ala arq GCA AGA	pro CCT	5 S	glu val GAG GTC	val GTC	ere CCA	leu CTT
20	245 cys TGC	leu CTG	ser leu ala TCA TTA GCT	ala GCA	asp GAT	phe III		val CTC	5 Q	thr
	gle GAA	lys A&A	ser TCA	tyr TAT	ala GCA	leu CTT	val GTA	3e 7	asn arg AAC AGG	57.4 50.5 100.
	thr ACG	ser	pro CCT	glu GAA	ala GCT	ale Gac	pro the leu val CCA ACT CTT GTA	leu CTA		asp 11e CAT ATA
	h Is CAC	ser 100	leu TTG	tyr Tat	ala OCC	392 cys TGT	thr	tyr Tat	leu val TTC CTC	asp
25	240 1ys val AAA GTC	270 ser 11e TCG ATC	asp GAC	leu TTG	361 cys 1CT	asn	pro CCA	450 glu asp GAA GAC	leu 11G	510 his ala CAT GCA
			300 ala CCT	330 phe TTT	360 cys TGC	390 gln CAA	420 thr ACT		480 ser TCC	510 h13 CAT
	thr	asp	50	met ATG	glu lys GAG AAG	lys AA	ser TCA	448 oys ala IGI CCA	28 8	phe TTC
	leu CTT	gln CAA	met ATG	91y GGC		11e ATC	val		thr ACA	thr
	thr asp ACA GAT	glu asn CAA AAT	g tu GAG	val phe leu GTC TTC TTG	leu CTA	gin asn leu CAG AAT TIA	g. CAA	ala lys arg met pro GCA AAA AGA ATG CCC	477 cys TGC	phe TTC
30	thr ACA		asp GAT	phe TTC	thr ACT	asn AAT	ور در	met ATG	476 cys TCC	thr ACA
	val GTG	265 cys TCT	val glu asn GTG GAA AAT	val GTC	thr ACC		val GTA	arg AGA	1ys AAA	g SA
	leu TTA	tyr lle TAT ATC	olu GAA	asp GAT	tyr glu IAT GAA	pro CCT	lys lys AAG AAA	1ys AAA	thr ACC	ala GCT
	ser lys leu val thr asp TCC AAG TTA GTG ACA CAT	265 lys tyr lle cys AAG TAT ATC TGT	glu val glu asn GAA GTG GAA AAT	321 glu ala lys asp val GAG GCA AAG GAT GTC	351 lys thr tyr glu thr thr leu AAG ACA TAT GAA ACC ACT CTA	glu glu pro GAA GAG CCT	411 tyr thr lys lys val pro TAC ACC AAG AAA GTA CCC	glu ala lys arg met pro GAA GCA AAA AGA ATG CCC	476 477 arg val thr lys cys cys AGA GTC ACC AAA TGC TGC	501 glu phe asn ala glu thr phe GAG TIT AAT GCI GAA ACA ITC
25		lys AAG	glu GAA	ele GCA	th ACA	91°	thr	91e	arg AGA	phe TTT
35	231 val GTT	261 ala GCC	291 ala GCC	321 glu GAG	351 lys AAG	381 val GTG	411 tyr TAC	14 170 170	471 asp CAC	201 94 86

	(1790)	570 phe ala glu qlu qly lys lys leu val ala ala ser gln ala ala leu qly leu ter TTT GCC CAG GAG GGT AAA AAA CTT GTT GCT GCA AGT CAA GCT GCC TTA GCC TTA TAA CATCACATTTAAAAG (1883)	ter ter ter ter (er
	င္တပ	ະ	A (2
	9 560 3 1ys r AAG	er Aaa	. ATT
5	558 559 cys cys TCC TCC	t. ATTT	1110
	8 55 G 75	TCAC	₹ ¥
	6 P	S S	AACA
	E 5	u te A TA	AAAA
10	558 559 pro lys ala thr lys glu gln leu lys ala val met asp asp phe ala ala phe val glu lys cys cys ccc aac cca aca aca aaa cac caa ctc aaa cct cat ca	y le	GTCT
10	do s TTT	- B B	Ž
	ا ع ج	e le	ACA
	ລຸ ຄວ	1 8 3 T	AGCC
	ro L	A 60	GTAA
15	550 9 88P	580 r 91n r C&A	76CT
	2 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	A AG	CGT
	I Ge	2 2	III
	ر در	ਵੌ <b>ਤ</b>	77.0
	2 S	- va	rerri
20	2 &	A CT	ATC
	A CT	8 <u>Y</u>	ratt(
		4 1 X	T.
	29	£ 8	AAA
	2.8	£ §	CATC
25	540 thr	570 91u	TCA/
	9 6 6	1 3 3 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	AAA
	Lys	Pho III	PAAAC
	200	567 978 10C	F 69.0
20	1ys AAG	glu thr GAG ACC	ter
30	M13 CAC	950	ter TCACA/
	1ys	1ys AAG	ACCA
	val lys	asp lys c	Tauti
	leu CTC	asp GAC	PTCAL
35	531 91u GAG	561 ala CCT	7707

ICATTITICCCICITITICICTGIGCTICAATIAAIAAAAAIGGAAAGAAICIAA..... 20 .....AA (2078)

Following are examples which illustrate procedures, including the best mode, for practicing the invention. These examples should not be construed as limiting. All percentages are by weight and all solvent mixture proportions are by volume unless otherwise noted.

#### 5 Example 1 Isolation of Messenger RNA

•

Human liver mRNA was obtained following the procedure of Chirgwin, et al [Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294-5299]. Immunoprecipitation of albumin containing polysomes was performed according to Taylor and Tse [Taylor, J.M. and Tse, T.P.H. (1976) J. Biol. Chem. 251, 7461-7467]. In vitro translation of mRNA was carried out in a reticulocyte cell-free system, following the instruction of the manufacturer (New England Nuclear). The translation products were separated electrophoretically according to Laemmli [Laemmli, J.K. (1970) Nature 227, 15 680-685.

#### Cloning Procedures Example 2

30

Double stranded cDNA was synthesized as described previously [Law, S., Tamaoki, T., Kreuzaler, F. and Dugaiczyk, A. (1980) Gene 10, 53-61]. It was annealed to PstI-linearized pBR322 DNA [Rolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crossa, J.H. and Falkow, S. (1977) Gene 2, 95-113] that had been tailed with 15 dG residues/3'-terminus [Dugaiczyk, A., Robberson, D.L. and Ullrich, A. (1980) Biochemistry 19, 5869-5873]. The annealed DNA was used to transform E. coli strain RR1, as detailed previously [Law, 25 S., et al., <a>Ibid.</a>]. The albumin clones were selected using the colony hybridization method of Grunstein and Hogness [Grunstein, M. and Hogness, D.S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961-3965], with [32p]-labeled cDNA synthesized with the immunoprecipitated polysomal mRNA as template.

As shown in Example 5, plasmids pHA36 and pHA206 were deposited in E. coli HB101 hosts. The plasmids were obtained from <u>E. coli</u> RR1 hosts, described in this example, and transformed into E. coli HR101 by standard procedures well known to those of ordinary skill in this The E. coli RR1 hosts were lysed and then centrifuged to art. 35 separate the chromosomal DNA, cell DNA and plasmid DNA. The plasmid DNA, remaining in the supernatant, is precipitated with ethanol and the precipitate is resuspended in buffer, e.g., TCM (10mM Tris.HCl, pH 8.0, 10 mM CaCl<sub>2</sub>, 10 mM MgCl<sub>2</sub>). The cells for transformation are

prepared as follows: 120 ml of L-broth (1% tryptone, 0.5% yeast extract, 0.5% NaCl) are inoculated with an 18 hour culture of HR101 NRRL B-11371 and grown to an optical density of 0.6 at 600 nm. Cells are washed in cold 100 mM NaCl and resuspended for 15 minutes in 20 ml chilled 50 mM CaCl<sub>2</sub>. Bacteria are then concentrated to one-tenth of this volume in CaCl<sub>2</sub> and mixed 2:1 (v:v) with annealed plasmid DNA, prepared as described above. After chilling the cell-DNA mixture for 15 minutes, it is heat shocked at 42°C for 2 minutes, then allowed to equilibrate at room temperature for ten minutes before addition of L-broth 10 times the volume of the cell-DNA suspension. Transformed cells are incubated in broth at 37°C for one hour before inoculating selective media (L-agar plus 10 µg/ml tetracycline) with 200 µl/plate. Plates are incubated at 37°C for 48 hours to allow the growth of transformants.

## 15 Example 3 Mapping of Restriction Endonuclease Sites

Restriction endonucleases were obtained from Rethesda Research Laboratories and New England Biolabs and were used according to the manufacturers' instructions. The digested DNA fragments were analyzed electrophoretically on agarose [Helling, R.B., Goodman, H.M. and Boyer, H.W. (1974) J. Virol. 14, 1235-1244] or acrylamide [Dingman, C., Fisher, M.P. and Kakefuda, T. (1972) Biochemistry 11, 1242-1250] gels.

### Example 4 DNA Sequencing

35

phosphatase (Worthington) and labeled at the 5'-ends with polynucleotide kinase (Boehringer-Mannheim) and Y[32p]ATP. Following digestion with a second restriction endonuclease and electrophoretic separation of the fragments, DNA sequence determination was done according to the procedure of Maxam and Gilbert [Maxam, A. and Gilbert, W. (1980) Methods Enzym. 65, 499-560] and the degradation products were separated electrophoretically on 0.4 mm acrylamide gels as described by Sanger and Coulson [Sanger, F. and Coulson, R. (1978) FEBS Letters 87, 107-110].

#### Example 5 Recombinant Plasmids pHA36 and pHA206

As disclosed in Example 2, albumin clones were selected by hybridizing to the enriched albumin cDNA probe. Plasmid pHA36 contained the largest insert of an albumin cDNA sequence. Both plasmids pHA36 and pHA206 have been deposited in a viable  $\underline{E}$ ,  $\underline{coli}$  host in the

permanent collection of the Northern Regional Research Laboratory (NRRL), U.S. Department of Agriculture, Peoria, Illinois, U.S.A. Their accession numbers in this repository are as follows:

HB101(pHA36) - NRRL B-12551

HB101(pHA206) - NRRL B-12550

5

 $\underline{\text{E. coli}}$  HB101 is a known and widely available host microbe. Its NRRL accession number is NRRL B-11371.

NRRL B-12550 and NRRL B-12551 are available to the public. upon the grant of a patent. It should be understood that the availability of these deposits does not constitute a license to practice the subject invention in derogation of patent rights granted with the subject instrument by governmental action.

<u>E. coli</u> RR1 and <u>E. coli</u> HB101 are known and widely available host microbes. Their NRRL accession numbers are NRRL B-12186 and NRRL 15
 B-11371, respectively.

pBR322 is a well known and widely available plasmid. It can be obtained from the following host deposit by standard procedures:

NRRL B-12014 - E. coli RR1 (pBR322).

YEp6 is a well known and widely available yeast episomal plasmid. 20 It can be obtained from the following host deposit by standard procedures:

E. coli HB101 (YEp6) - NRRL B-12093.

# Example 6 Assembly of the Serum Albumin Gene

Assembling the pieces together is a straighforward task of restriction enzymology. There is only one MspI site in the overlapping
DNA sequence of the two cDNA clones. Two enzymatic steps of (i) MspI
digestion of the two DNAs, followed by (ii) the use of ligase, an
enzyme that seals DNA fragments, will give the desired product.
Although two other undesired DNA species will also be obtained in the
course of this recombination reaction, both of them will differ substantially in size. Thus, separation and isolation of the desired DNA
species will be achieved.

The assembled DNA clone can be used to transform two types of cells:

(a) Escherichia coli

- (b) Saccharomyces cerevisiae
- (a) The vector of choice is plasmid pBR322, the same that has

been successfully used for cloning of the two fragmented pieces of the serum albumin cDNA.

(b) In order to transform yeast with the serum albumin structural gene sequence, the DNA must be inserted into one of the 5 existing yeast plasmid vectors. This can be accomplished by taking advantage of the fact that several restriction endonuclease recognition sequences are absent from the cloned serum albumin DNA. Synthetic EcoR1 DNA linkers can be ligated to the DNA fragment containing the serum albumin sequence followed by insertion (ligation) into one 10 of the yeast plasmid vectors, e.g., YEp6, at the Eco R1 cloning site. The fused chimeric plasmid can be used to transform yeast according to an established procedure [Hinnen, A., Hicks, J.B. and Fink, G.R. (1978) Proc. Natl. Acad. Sci. USA, 75, 1929]. YEp6 can be obtained from the NRRL repository, as disclosed supra.

# 15 Example 7 Expression of the Serum Albumin Gene

The main body of the structural gene will be transcribed by the E. coli or yeast enzymes. If little or no albumin is produced with the selected host, then an Escherichia coli promoter DNA sequence carrying an initiation codon, i.e., ATG, can be ligated at the begin-20 ning of the serum albumin structural gene. Such elements are known and available, e.g., lac promoter used for the expression of human interferon gene in E. coli [Proc. Natl. Acad. Sci. 77, 5230 (1980)]; source of promoter DNA [Proc. Natl. Acad. Sci. 76, 760 (1979)]. Also, see Nature, Vol. 281, October 18, 1979. It has already been 25 documented that such Escherichia coli promoter sequences function well in the expression of foreign genes in Escherichia coli [Mercereau-Puijalon, O., Royal, A., Cami, B., Garapin, A., Krust, A., Gannon, I. and Kourilsky, P. (1978) Nature 275, 505; and Goeddel, D.V., Kleid, D.G., Bolivar, F., Heyneker, H.L., Yansura, D.G., Grea, R., Hirose, 30 T., Kraszewski, A., Itakura, K., and Riggs, A. (1979) Natl. Acad. Sci. USA 76, 106]. For expression in yeast, see Rose, M., Casadaban, M.J. and Botstein, D. (1981) Proc. Natl. Acad. Sci. USA 78, 2460 and 4466. Screening of Clones Producing Albumin Example 8

Immunological methods can be used to detect small amounts of albumin made in a bacterium. Flat disks of flexible polyvinyl are coated with the IgG fraction from an immune serum and the disks are pressed onto an agar plate so that antigen released from an <u>in situ</u> lysed microbial colony can bind to the fixed antibody. The plastic

disk is then incubated with the same total IgG fraction labeled with radioactive iodine so that other determinants on the bound antigen can in turn bind the iodinated antibody. Radioactive areas on the disk expose X-ray film during autoradiography and thus identify colonies producing the protein which is being screened for. Detailed protocols of this procedure have been published [Broome, S. and Gilbert, W. (1978) Proc. Natl. Acad. Sci. USA, 75, 2746]. The purification of human serum albumin can be accomplished by using procedures well known in the art. For example, procedures disclosed in a chapter by T. Peters: Purification and Properties of Serum Albumin, in: The Plasma Proteins, Putnam, Ed. Academic Press, New York, 1975, can be used.

The work described herein was all done in conformity with physical and biological containment requirements specified in the NIH Guidelines.

## CLAIMS

- 1. Plasmid pHA36, having a restriction endonuclease pattern as shown in the drawing.
- 5
  2. Plasmid pHA206, having a restriction endonuclease pattern as shown in the drawing.
- 3. E. coli HB101 (pHA36) having the deposit accession number 10 NRRL B-12551.
  - 4. E. coli HB101 (pHA206) having the deposit accession number NRRL B-12550.
- 5. A microorganism modified to contain a nucleotide sequence coding for the amino acid sequence of human serum albumin; said nucleotide sequence is as follows:

20

25

	(30)	(170)	(092)	(350)	(084)	(330)	(620)	(710)	(300)
	Ser ACC	20 179 AA	50 818 GCA	8 2 E	110 CCA	140 try TAT	170 gln CAA	200 Cys	230 glu GAA
5	ohe TTT	phe TTC	phe TTF	thr	asa	leu TTA	169 cys TCC	1ys AAG	<del>ه ا ه</del> ۵۲۸
)	lau CTC	AAT	ale CAA	8 to GCA	ese GAC	tyr TAC	168 cys 1GT	leu CTC	phe TTT
	phe	915 GA	thr ACT	val GTT	asp	1ys AAA	<b>5</b> 5	ara Aga	a)u GAG
	lau CTT	5 AS		thr ACA	1 ys Aaa	lys AAA	rhr ACA	aln CAG	ala CCT
	9	<del>2</del> 8	₹.89	55 57 757	h Is CAC	leu 11G	phe TTT	lys AA	1ys AAA
10		leu TTG	asn AAT	leu TTA	ca Ca	glu glu thr phe leu GAA GAG ACA TTT TTG	ala GCT	ချေ	<b>5</b> 22
		as <b>p</b> GAT		1ys AAA	leu TTG	& hr	ala CCT	ser ICT	phe TTT
		lys AAA	leu TTA	asp GAC	phe TTC	a)u GAG	1ys AAA	ser TCG	gln arg CAG AGA
		phe TTT	1 ys	91.y CGA	100 101 q1u cys CAA TCC	5 & S	tyr Tat	ele GCT	gln
		10 his arg CAT CGG	40 val GTA	70 phe TTT	55 ± 89	130 asn AAT	160 arg AGC	190 91y lys CGC AAG	220 3er AGC
15		h I s CAT	h1s CAT	leu	AAT	esp GAC	1 ys	91y 666	leu CTG
		ele GCT	asp GAT	thr ACC	ACA ACA	his CAT	ala GCT	5 88 88	ဦး ၁၃၁
		val	glu GAA	A Is	g 1 y	phe TTT	ohe TTT	esp GAT	ala GCT
		glu GAG	phe	leu CTT	873 CCT	ala GCT	phe TTC	9 9 9 9 9	phe lys ala trp ala val TTC AAA GCA TGG GCA GTA
20		ser	g S	ser TCA	ale GAA	th. ACT	leu CTT	leu	e1.8 A
		lys AAG	¥ 25	lys AAA	41 CAA	124 039 100	leu CTC	gle GAA	4 rp
		n13 Cac	gln CAG	asp GAC	1 ys	met ATG	g Ju GAA	asp CA1	25 CCA
		ala GCA	gln CAG	62 cys TGT	91.9 6CA	val GTG	Pro CCG	lys leu AAG CTC	210 ala phe lys GCT TTC AAA
		asp CAT	leu CTT	60 glu asn GAA AAT	91 6ys TGT	esp	၁ <b>၂</b> ၈ ဝင်င	lys AAG	phe TTC
25		- 25 25 25 25 25 25 25 25 25 25 25 25 25 2	30 tyr tat		8 8 55	120 glu val GAG GTT	150 tyr TAT	160 Pro CCA	210 e1a GCT
		r o phe arg	gla CAG	ala GCT	asp GAC	glu GAG	phe TTT	leu leu CTG TTG	P AS
		r o phe	ala GCT	ser TCA	್ಟಿ ೧೧	5 52 525	eyr TAC	leu CTG	95 85
		P val CTG	phe TTT	glu GAG	met ATG	arg AGA	arg his pro	177 ala cys GCC TGC	91.9 CCA
20		p gly val GCT CTC	ala 000	esp	glu met GAA ATC	i arg leu val arg CGA TTG GTG AGA	hts CAT	e 1 e	phe III
30		erg ACC	11e ATT	ala GCT	gly ccr	leu TTG	arg Aga	ele	<b>1</b>
		tyr ser	leu TTG	val CTT	tyr Tat	5 45 55	P. S.	1ys AAA	<b>క్</b> క
		er ala tyr ser TCC GCT TAT TCC	21 ala leu val leu ile GCC TTG GTG TTG ATT	53 thr oys val ala ACA IGT GTT GCT	glu thr GAA ACC	5 S	- ag 3	esp CAT	210 ser leu gin lys phe gly glu arg ala AGT CTC CAA AAA TIT GGA GAA AGA GCT
		ala CCT	leu TTG	<b>thr</b> ACA		111 asn leu AAC CTC	11e att	ele CCT	ser AGT
35		se <b>r</b> 100	21 ala GCC	12 sys	arg CCT	111 asn AAC	ale SA	171 e1e CCT	20 818 878

	(890)	(086)	(1070)	(1160)	(1250)	(1340)	(1630)	(1520)	(1610)	(1700)
5	260 leu CTT	290 11e ATT	320 ala GCT	320 ele CCC	380 leu CCT	& 10 8 T Q CCT	#40 N1s	\$70 ser AGT	Sno 1ys AAA (	530 val CTT (
	BSP	289 cys TGC	eyr TAT	leu CTT	pro CCT	val GTT	173 AAA		200	
	8	h í s CAC	asn AAC	ara Aga	1 ys AAA	leu TTA	438 cys TGT	pro val CCA GTA	val p	ala leu GCA CIT
:	org AGG	ser TCC	1 ys		phe TTT	leu CTG	437 cys TGT		tyr v TAC G	
	esp GAC	1 ys	316 0ys 1GC	Jeu CTG	glu phe CAA TTT	# B	lys AAA	lys the Aaa acg	thr t ACA T	oln thr CAA ACT
10	esp CAT	glu GAA	316 val oys lys CTT TCC AAA	leu CTG	asp CAT	AAT	ser Acc /	alu GAG A	asp glu thr tyr val GAT GAA ACA TAC GTT	8 X8
	253 cys 218 1GT GCT	leu 776	asp	v81 GTG	phe	GAG A	41¢	his q	asp q	lys lys aag aaa
	253 cys 1GT	leu leu CTG TTG	lys AAG	val GTC	val GTG	phe 1	val GTG (	leu his TTG CAT		16 J
	leu glu CTT GAA	pro CCT	ser ACT	ser val val leu leu leu TCT GTC GTG CTG CTG	TYS AAA	1ys AAA	lys val gly AAA GTG GGC	vel l	glu val CAA GTC	- A
15	250 leu leu CTG CTT	ay A	28 88	tyr	ala SCC		41. 80.0	461 6ys v TGT (	leu g CTG C	2 g
15	250 leu CTG		310 val CTT	340 asp CAT	370 tyr TAT	\$00 glu kyr GAG TAC	430 leu qly CTA GGA	460 4 leu c TTA 1	490 ala leu GCT CTG	520 glu arg ain 11e 1ya 1ya GAG AGA CAA ATC AAG AAA
	asp CAT	279 cys TCT	phe TTT	pro CCT	369 Cys TCC	F 55	asn AAC	gln J CAG J		
	919 664	278 cys TCC	asp GAT	hIs CAT	S & A	leu CTT	AGA AGA	asa g	phe s	glu lys CAG AAG
	h le	5 A	ala GCT	879 AGG	h s CAT	al CAG	ser TCA	leu a	ر ا ا	ser q
20	245 246 Oys cys his TGC TGC CAT	ser lys leu lys AGT AAA CTG AAG	ala GCT	arg AGA	pro CCT	gle GAG		val GTC	pro cys phe ser CCA TGC TTT TCA	leu s
		leu C7G	leu TTA	ala GCA	asp CAT	phe TTT	glu val GAG GTC	val GFG	919 CGA	514 cys thr leu TGC ACA CTT
	g alu	1ys AAA	ser TCA	glu tyr GAA TAT	ala GCA	leu CTT	val GTA	ser 700	arg	514 cys 1CC
	val his thr glu GTC CAC ACG GAA		pro CCT	glu GA	ala CCT	ale GAC	leu CTT	leu CTA	asn AAC	
	F. CAC	ser TCC	leu TTG	tyr TAT	ala တင	392 cys TGT	thr ACT	tyr	val GTG	asp 11e Cat Ata
25		270 ser 11e TCG ATC	asp GAC	Jeu TTG	361 0ys TGT	asn AAT	pro CCA	asp GAC	leu TTG	ele GCA
			300 a1a ccT	330 phe TTT	360 993 160	390 gln CA	420 ehr ACT	450 910 GAA	480 ser leu TCC TTG	510 hla CAT
	thr ACC	asp GAT	pro CC1	met ATG	lys AAG	1ys AAA	ser TCA	818 GCA	glu GAA	phe 1
	leu CT?	g SA	met ATG	phe leu gly TTC TTG GGC	glu thr thr leu glu GAA ACC ACT CTA GAG	11e ATC	val GTG	448 0ys TCT		
30	asp CAT	AAT	g tu Gac	leu 17G	leu glu CTA GA	leu TTA	eg Pr	ရှင် သ	477 cys TGC	phe thr TTC ACC
	the S ACA	g Ju	asp		thr	asn	. წეე	met ATG	476 0ys TGC	- <u>-</u> 5
	leu val the asp le 17A GTG ACA GAT CT	269 cys TGT	giu val glu asn GAA GTG GAA AAT	asp val GAT GTC	thr	glu glu pro gin asn leu ile GAA GAC CCT CAC AAT ITA ATC		arg AGA	476 477 thr lys cys cys thr ACC AAA TGC TGC ACA	glu thr phe thr GAA ACA TTC ACC
	leu ? T.A	11e ATC	glu GA	asp GAT	91¢ 684	pro CCT	1ys AAA	1ys AAA	thr Acc	ala GCT
	ser lys ICC AAG	eyr Tat	val CTG	ala lys GCA AAG	tyr TAT	g lu GåG	1ys AAG	ele GCA	val GTC	AAT
35		AAG	<b>3</b> € .		thr	91c GA	thr	CA S	arg val AGA GTC	phe asn ala glu thr III AAT GCT GAA ACA
	231 val GTT	261 ala ccc	291 ala GCC	321 91u GAG	351 1ys AAG	381 val CTC	411 tyr TAC	441 pro	471 25p	501 g1u GAC

	(1790)	570 glu glu qly lys lys leu val ala ala ser gln ala ala leu qly leu ter GAG GAG GGT AAA AAA CTT GTT GCT GCA AGT CAA GCT GCC TTA GCC TTA TAA CATCACATTTAAAAG (1883)	MGAAAATGAAGATCAAAAGCTTATTCATCTGTTTTTCTTTTTCGTTGGTGTAAAGCCAACACCCTGTCTAAAAACATAAATTTCTTTAA (2002)
	2	Ē	(2)
	559 560 eys lys TCf. AAG	, r	¥Ε
5	55.9 100	2 1	1161
	538 638 130	25	AAA
_	1y3	CAT	ACAT
	200	ž Ę	WW.
	<b>5</b> 5 1	leu TTA	ſζŢ
10	ag E	<del>,</del> 5	212
	als OCT	3 E	NC ACC
	s1s GCT	<b>*</b> 55	<b>₹</b>
	phe TTC	ala GCT	Ž.
	55 g 57	580 gln CAA	SCTG1
15	\$ 2	ser	) 20110
	met ATG	31.8 GCA	1116
	5 75	ala GCT	TCTI
	<b>5</b> 50 15	val GTT	HI
20	<b>₹</b>	Jeu CT	VTCTG
	1e2 CTC	1ys AA	TTC.
	E YS	1 X X	CTTA
	95 646	91y GGT	<b>₩</b>
	<b>\$</b> ₹	9 20	ATCA
25	550  1ys ala thr lys glu gln leu lys ala val met asp asp phe ala ala phe val glu lys cys and can ach	570 glu GAG	8
	# 55 # 55	phe ala TTT GCC	₩
	lys AG	ag I	<b>ઇ</b> ₹
	<b>2</b> 23	267 153 153	<b>ઇ</b> ઇ
20	lys AG	glu thr GAG ACC	ter ter ITCACAATAA(
30	• ¥2 C} €	95 66 66	ઇ
	<u>*</u> . ₹	lys AG	CCAT
	val lys his lys GIG AA CAC AAG	asp lys	CCTA
	Leu	asp GAC	īc A Ģ
35	91c GAG	561 ala CCT (	ter ter Catctcagcctaccatgagaataagaa

TCATTTTGCCTCTTTTGTCTGTGCTTCAATTAATAAAATGGAAAGAATCTAA..... 20 .....AA (2078)

6. Nucleotide sequence of the cDNA of human serum albumin, said nucleotide sequence is as follows:

5	(170)	(192)	(350)	(440)	(330)	(620)	(710)	300)
	20 1ys AAA	50 ala GCA	80 Leu CTT	110 Pro	140 try TAT	170 qln CAA	200 0ys TGT	230 glu GAA (300)
	phe TTC	phe TTT	thr	asn	leu TTA	169 707	1ys AAG	ala GCA
	asn	thr alu phe ACT CAA TTT	ala thr GCA ACT	ase GAC	tyr	168 cys TGT	leu CTC	phe TTT
10	o Ay	th <b>r</b> ACT	val GTT	asp GAT	1ys AAA	25	ara Aga	ale CAG
10	91. GA9	val GTA	thr ACA		1ys	chr	aln CAG	ala CCT
	lys asp leu aly alu alu asn phe AAA GAT TTG GGA GAA AAT TTC	40 his val lys leu val asn glu val thr Cat gta aaa tta gtg aat gaa gta act	55 155	phe leu gin his lys TTC TTG CAA CAC AAA	glu glu thr phe leu lys lys tyr Gaa gag aca tty ttg aaa aaa tac	phe	1ys	1ys
	Jeu TTG	40 val lys leu val asn GTA AAA TTA GTG AAT	75 lys leu cys AAA TTA TGC	c d a	phe TTT	ala phe CCT TTT	91a	pro CCC
	lys asp AAA CAT	val GTG	asp lys	1ec 77G	thr ACA	e)a GCT	ser TCT	phe
15	1ys AAA	leu TTA	98 <b>9</b> CAC	phe TTC	alu GAG	1ys	ser TCG	ACA ACA
	phe	lys AAA	<del>,</del> 8	101 103 100	ale GAA	tyr TAT	ala CCT	gln arq CAG AGA
	10 819 000	. 40 val	70 phe 117	ale GAA	130 asn AAT	160 erg tyr lys ele ele ACC TAT AAA GCT GCT	190 1ys	
	elo hís GCT CAT	ZA T	thr leu ACC CTT	asn	asp GAC	1ys AAA	<del>7</del> 55	leu CTG
00		phe alu asp his TTT GAA CAT CAT	thr	8.78 ASA	hls CAT	ata CCT		ဦ ည
20	glu val GAG GTT	<u>4</u>	bis Cat	<del>,</del> 55	ala phe GCT TTT	ohe TTT	asp GAT	ala GCT
		phe TTT	leu CTT	Pro CCT		phe TTC	550 550	
	ser AGT	or A	ser TCA	glu GAA	thr ACT	leu	leu CTT	trp ela vel TGG GCA GTA
	1ys AAG	34 0ys 1CT	1ys AAA	g]a CAA	12% 0y3 TGC	leu CTC	ale GAS	trp ala TGG GCA
25	h1s CAC	gln CAG	asp GAC	1ys AAA	met ATG	glu	leu asp CTC GAT	lys alæ AAA GCA
	စ <b>ါ</b> စ <b>GC</b> A	gln CAG	62 0ys 1G1	ala GCA	val GTG	970	leu CTC	1 ys
	sap GAT	leu CTT	60 glu asn GAA AAT	93 93 1GT	asp	618 CC	lys AAG	phe TTC
		30 ala gin tyr leu GCT CAG TAT CTT		90 cys TGC	120 val GTT	150 tyr TAT	180 Pro CCA	210 ala GCT
		ala gin CCT CAG	ela GCT	asp CAC	g r GAG	tyr phe TAC TTT	leu leu. 27G 77G	arg AGA
30		ala GCT	38. 1CA	ala CCT	pro CCA	tyr TAC	leu CTG	<b>28</b>
		phe TTT	glu GAG	glu met GAA ATG	arg AGA	pro CCT	177 933 TCC	91. 86.4
		a1a CCC	val ele esp GTT GCT GAT		val GTG	ele arg erg his pro CCC ACA ACA CAT CCT	် ရေး ဗေ	lys phe AAA TTT
		11e ATT	ala CCT	gly GGT	Jeu 77G	arg ASA	ala GCT	1ys AA
35		leu TTG	val GTT	tyr TAT	ere SSS	arg ASA	lys AAA	eg CAA
		leu val leu ile ala phe TTG GTG TTG ATT GCC TTT	51 53 lys the cys val AAA ACA TGT GTT	glu thr GAA ACC	leu pro arg leu val CTC CCC CGA TTG GTG	ala 800	177 asp lys ale ale oys GAT AAA GCT GCC TGC	leu gln CTC CAA
		leu TTG	thr ACA		leu CTC	11e Att	ala CCT	ser AGT
		21 ala CCC	173 A&A	81 arg CGT	111 asn AAC	141 glu GAA	171 818 50T	201 818 CCC

5	(890)	(980)	(1070)	(1160)	(1250)	(1340)	(1430)	(1520)	(1610)	(1700)
	260 leu CTT	290 11e ATT	320 ala CCT	350 a1a ccc	380 leu CCT	\$10 879 CGT	440 his cat (	&70 9er. Agt (	500 1ys AAA (	530 va) CTT (
	asp CAC	289 cys TCC	tyr TAT	le CTT	67 CCT	val GTT	lys P		pro 1	
	arg ala AGG GGG	N 18 CAC	asn	5 Q	1ys		438 cys   TCT /	oro val CCA GTA	val p	ala leu CCA CTT
10	A SC	ser Icr	lys Aaa	leu CTG	phe ITT	leu leu CTG TTA	437 cys TGT		tyr v	
	asp asp CAT CAC	glu lys GAA AAA	316 cys rcc	leu CTG	g]u CAA	# S	lys c	lys thr AAA ACG		aln thr CAA ACT
	ass CAT	g]u GAA	val GTT	leu CTG	asp CAT	AAT	ser Acc	alu J CAC A	alu thr GAA ACA	ys g AA C
	ala CCT	pro leu leu CCT CTG TTG	asp CAT	val GIG	phe	c Ac	4. 500 500	his e	ess CAT C	gin ile iys iys CAA ATC AAG AAA
	253 cys TGT	Jeu CTG	lys AAG	val GTC		phe TTC	val GTG	leu h		le 1 TC A
15	leu glu CTT CAA		ser AGT	ser	lys val AAA GTG	1ys AAA	1ys y	val 1 GTG 1	glu val CAA GTC	1 A A
	leu CTT	1ys AAA	glu GAA	ty. TAC	ီး ဗင	TAC	414 665		leu g CTG C	
	250 asp leu CAT CTG	280 91u CAA	310 val GTT	340 83p CAT	370 tyr TAT	400 910 CAC	430 leu qly lys val CTA GCA AAA GTG	&60 &61 leu cys TTA IGT	490 ala 1 GCT C	520 glu arq GAG AGA
		279 cys TGT	phe TTT	pro CCT		4 8 8	asn 1			
20	his gly Cat gga	278 cys TGC	889 CAT	h is CAT	369 alu cys CAA TGC	Jeu CTT	P. S. S.	leu asn gln CTG AAC CAG	phe ser ITT ICA	qlu lys CAG AAG
20		of A GA	ala CCT	arg AGG	h1s CAT	CAG	ser arg TCA AGA	leu a		ser q TCT G
	246 cys 1GC	leu lys CTG AAG	ala CCT	arg AGA	pro CCT	a tu		val cor	pro cys CCA TGC	leu ser CTT TCT
	2&5 cys 1CC		leu TTA	ela GCA	asp CAT	phe TTT	glu vel CAG GTC	val v	5 43 6 43	514 cys thr leu TCC ACA CTT
	thr glu ACG GAA	ser lys AGT AAA	ser TCA	tyr TAT	ala GCA			ser v	arg a	514 cys t TCC A
25	th. ACG	ser	pro	glu tyr CAA TAT	ala CCT	glu leu GAG CTT	leu val CTT GTA	leu s CTA 1	asn a	
	h Is CAC	ger TCC	leu TTG	tyr TAT	ال 200	392 cys TCT	the 1	tyr 1 TAT C	val a	asp Ile Cat Ata
	val GTC	ile ATC	asp GAC	leu TTG	361 cys TGT	AAT	Pro CCA	839 GAC 1	leu v TTG G	ala a CCA G
	240 173 AAA	270 ser TCG	300 ala cci	330 phe TTT	360 cys 160	390 91n CAA	\$20 thr g	\$50 glu a	480 ser leu 7CC TTG	510 h1s a cat c
	thr ACC	asp CAT	pro CCT		1 ys AAG		ser (	ala g CCA C	glu s	5 phe h TTC C
30	leu CTT	gln CAA	met AT <b>G</b>	gly met CCC ATG	glu	Ile lys ATC AAA	val s	448 0ys 8	thr g	thr p ACC I
	asp CAT	glu asn CAA AAT	g)u GAG		leu CTA		gla c	pro c	73 73 75 75	
	th <b>r</b> ACA	glu GAA	asp glu GAT GAG	phe 7TC	thr leu ACT CTA	AAT	. 579 CCC C	met p	476 477 cys cys TGC TGC	4 2
	ser lys leu val thr asp TCC AAG TTA GTG ACA GAT	265 cys TGT	glu val glu asn asp glu GAA GTG GAA AAT GAT GAG	val GTC	thr	18 S		arg met AGA ATG	476 477 1ys cys cys AAA TGC TGC	glu thr phe GAA ACA TTC
35	leu TTA	tyr 11e TAT ATC	glu val glu asn GAA GTG GAA AAT	asp CAT	glu thr GAA ACC	50.	8	8	thr 1 ACC A	
33	1ys AAG	tyr IAT	val GTG	1ys AAG		15 S	lys lys Aag aaa	glu ala lys GAA CCA AAA	val t	phe asn ala TTT AAT GCT
	ser TCC	lys tyr AAG TAT	g g GA	ala SCA	thr	28	the 1 ACC A	glu a CAA C	arg v	he a: TAA
	231 val CTT	261 ala lys GCC AAG	291 ala ccc	321 glu ala lys asp val phe leu GAG GCA AAG GAT GTC TTG	351 lys thr tyr AAG ACA TAT	JB1 val glu glu pro gln asn leu GFG GAA GAG CCT CAG AAT TFA	411 tyr thr lys lys val TAC ACC AAG AAA GTA	441 Pro 9 CCT G	471 asp arg val thr GAC AGA GTC ACC	501 glu phe CAG TTT
							•			J. 5. C

		_	_	
5	(1790)	(1883)	ter ter carcetaccataraaaaaaaaaaaaaaaaaaaaaaaaaaa	
	560 178 AAG	AG	TAA	
	85 8 8 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9	570 phe ala glu glu gly lys lys leu val ala ala ser gln ala ala leu qly leu ter TTT GCC CAG GAC GGT AAA AAA CTT GTT GCT GCA AGT CAA GCT GCC TTA GCC TTA TAA CATCACATTTAAAAG	TETT	
	558 5 1ys cys c AAG TCC 1	ACA1	AAATI	
10	1ys AAG	CATC	ACAT/	
10	vel glu l GTA GAG	ter TAA	AAAA	
	vel GTA	leu TTA	TCTA	
	550 glu gin leu lys els val cet asp asp phe els els phe e GAG CAA CTG AAA GCT GTT ATG GAT GAT TTC GCT RCT TTT (	<del>2</del> 88	ວເວເ	
	010 GC1	leu TTA	VACAC	
15	919	ala 6	/OCC/	
	phe TTC	65.5	GTAA	
	350 88p	580 r gln T CAA	1661	(8)
	8 8 8 8	a se A AG	TCGT	02) 1
20	i ae	1 e 1	:YTT	\$
		2 E	rrr	
	ີ 6 ອ ຮຸ≱	2 € 2 €	CTGT	<u>ب</u>
	26 75 75	\$ \$ \$	TCAT	A1
	H A	ys P	TTAT	\rcr
25	Pag o	11, 1 10, 1	VAAG	AGA
	173 84 9	alu c	ATCA	TGGA
	\$40 ala thr lys g cca Aca Aaa G	91u glu	GAAG	AAAA
	lys ala	۳ <del>۱</del> ۵	MAAT	WTAA
30	Lys AG	phe TTT	VAAG	4ATT/
	pro	567 cys 100	4G6	crtc
	Pys AG	thr ACC	ter ter ITCACAATAA	TCTG
	ris CAC	glu GAG	ter TCAG	7.01
35	. AS	1ys	ACCA	ICTT
Jü	val crc	567 s asp asp lys glu thr cys r CAC GAT AAG CAG ACC TCC	\CC1	10001
	531 giu leu val lys his lys r CAG CTC GTG AAA CAC AAG (	E S	rcTC/	TCATTITGCCTCTTTTCTCTGTGCATTAATAAAAATGGAAAGAATCTAA 20AA (2078)
	8 4 8	561 ala GCT	ક	5

7. Nucleotide sequence coding for the prepeptide of human serum albumin, said nucleotide sequence is as follows:

		- 4	• •	43
		(30)		
5		AGC		
		å E		
		leu bhe CTC TTT		
	·	-10 ser leu leu phe l TCC CTT CTT TTT C		
10		2 E		
10		5 E -		
		36T TCC		
		11e ATT		_
		phe 11e		
15		0 5 0		
		7 42 CTA		
		<del>1</del> 55		
		1ys AG		
	•	ATG ATG		
20		<b>S</b> Y		
		-18 p r Het 1ys trp val tl TTTCTCTTCTGTCAACCCACAGCCTTTGGCACA ATG ANG TGG GTA AC		
		נננו		
		ACAG		
25		ວວວ		
		ŢÇ <b>Ą</b>		
		101	- 1	క్ర
		'CTCI	r o -1 phe arg arg	CCT
		H	٦ ۾	H
30		8	۳ <u>د</u>	CIC
			-6 pr argglyval	55
			Ť .	Y
		ı	- i	ក្ត
35			tyr	<b>I</b> T
			=======================================	8
			200	55

8. Nucleotide sequence coding for pro human serum albumin, said nucleotide sequence is as follows:

5	(170)	(192)	(350)	(940)	(330)	(420)	(017)	(300)
	20 17s	50 818 GCA	80 leu CTT	110 pro CCA	140 try TAT	170 91n CAA	200 cys 1GT	230 glu GAA
	phe TTC	phe TTT	thr	asn	leu TTA	169 673 100	178 AAG	ala GCA
		of a	ela GCA	aso GAC	tyr TAC	168 cys TGT	teu CTC	phe TTT
10	qlu qlu gsn GAA GAA AAT	thr ACT	val	asb CAT	1ys AAA	<b>3</b> 8	P S S	alu GAG
10	g to	val	th <b>r</b> ACA	1ys AAA	1 ys	thr	ala CAG	ala CCT
	asp leu aly c	gg gg		his CAC	leu 76	phe TTT	1 <b>y</b> 3	1 ys
	leu TTG	asn	leu TTA	gln CAA	phe	ala phe GCT TFT	ala lys CCC AAA	pro CCC
	esp	val GTG	1 ys	leu 17G	AG #	ala GCT	ser TCT	ohe TTT
15	1ys AAA	leu val TTA GTG		phe TTC	. P & S	Nys AAA	ser ser TCG TCT	arg AGA
	phe FFT	1ys AAA	5 45 5 45 5 45	101 673 160	916 GAA	tyr TAT		gln CAG
	ე გ.გ ე	40 val GTA	70 phe TTT	5 5 8 6 4 8	130 asn AAT	160 arq AGG	190 173 AGC	220 ser AGC
	his CAT	E S	leu CTT	asn AAT	asp GAC	1ys AAA	<del>,</del> 88	leu CTG
	ola GCT	asp CAT	his thr	P P P	h s	ala GCT	a]e G∳	518 CC
20	val CTT	<b>2</b> 89	his CAT	91y GGG	phe	ohe TTT	asp CAT	ela CCT
	ser glu AGT GAG	phe TTT	ser leu TCA CTT	pro CCT	al a CCT	leu phë CTT TTC	5.00 000	val
	ser AGT	S S S	ser TCA	glu GAA	thr	leu CTT	leu	ala CCA
	1ys AAG	34 0ys 1GT	asp lys GAC AAA	gIn CAA	124 cys 760	leu CTC	g). GA	t.ը 166
25	h is CAC	gln CAG	asp	1ys AAA	met ATG	ag	asp GAT	ala CCA
~~	910 GCA	gla	62 0ys 1GT	ala GCA	val GTG	ဥ္သ	leu CTC	lys AAA
	esp CAT	leu CTT	asn	93 93 1GT	asp GAT	ala OCC	1ys AAG	phe TTC
	- 58	30 tyr TAT	60 91u GAA	90 993 TGC	120 val GTT	150 tyr TAT	180 Pro CCA	210 ala GCT
	erg CGT	gln CAG	ala CCT	asp GAC	g)u GAG	phe TTT	leu TTG	arg AGA
30	phe TTT	ala CCT	ser TCA	ala CCT	5 5 5 5 5	tyr TAC	leu CTG	5 8 8
	و د د د د	phe TTT	g lu GAG	met ATG	AGA	pro	177 978 1GC	
	91,4	ala ကေ	asp CAT	glu met GAA ATG	i leu val arg TTG GTG AGA	his pro Cat cct	a1 a	phe gly TTT CCA
	arg ACC	11e ATT	ala GCT	gly GGT	leu 77G	P 2	ala CCT	gin lys CAA AAA
		leu TTG	val	tyr TAT	5 83 83	8 4 8 4	1ys AAA	gln
35		val GTG	53 0ys 1GT	thr		ala CCC	GAT	leu CTC
		leu 1TG	thr ACA	glu GAA	leu CTC	11e ATT	ela GCT	Ser
		21 ala GCC	173 A&A	81 arg CGT	111 asn AAC	£ 28	171 ala cct	20 10 10 10 10 10 10 10 10 10 10 10 10 10

5	(890)	(980)	(1070)	(1160)	. (1250)	(1340)	(1430)	(1520)	(1610)	514 cys thr leu ser glu lys glu arg gln ile lys lys gln thr ala leu val TGC ACA CTT TCT GAG AAG GAG AGA ATC AAG AAA CAA ACT GCA CTT GTT (1700)
	260 1eu CTT	290 11e ATT	320 ala ect	350 ala ccc	383 CCT	& 10 arg	440 h13 CAT	&70 ser AGT (	500 I ys AAA (	530 val CTT (
•	ola asp ccc cac	289 Cys 10C	tyr Tat		878 CCT	val CTT	lys P	_	Pro 1 CCC A	3 € ~ 3 € ~
	ele SCC	h I v CAC	asn	arg ACA	1 ys	leu TTA	438 cys 1	<b>Dro val</b> CCA GTA	val p CTT C	~ 5 ~ 5
10	P 25	ser TCC		leu CTG	phe TTT	leu CTG	437 438 cys cys TCT TCT	thr p	tyr «	# 5
	485 GA 78	1 ys AAA	316 cys lys TCC AAA	leu CTG	g alu			lys t AAA A		= = = = = = = = = = = = = = = = = = =
	ele osp esp org CCT CAT CAC ACC	pro leu leu glu lys CCT CTG TTG GAA AAA	316 ser lys asp val cys lys act aac cat cit icc aaa	ser val val leu leu leu arg leu TCT GTC GTG CTG CTG ACA CTT	ess CAT	asm ala AAT CCC	ser lys ACC AAA	alu lys CAC AAA	alu thr GAA ACA	2 X
		leu TTG	88 <b>5</b> CAT	val		g ju	*	is a	asp a	7. S. 2.
	glu cys	leu CTG	Tys AAG	val GTC	val phe GTG TTC	phe o	val aly GTG GGC	leu hís TTG CAT		Ic 1 <sub>1</sub>
15		pro CCT	ser A <b>GT</b>	ser TCT	ala Iya val phe GCC AAA GTG TTC	Lys AAA	636 Leu qly lys val qly CTA GGA AAA GTG GGC	val l	glu val CAA GTC	= A
	250 leu leu CTG CTT	PA A	a) Ge		ala lys CCC AAA	tyr lys Tac aaa	419 665 4	461 Cys v TGT G	leu g CTG G	6 S
		84 288 84 288	310 val CTT	340 asp tyr CAT TAC	370 tyr	400 glu tyr CAG TAC	436 lev c CTA (	\$60 & leu c TTA T	490 ala leu GCT CTG	520 91u a
	gly asp GCA CAT	279 cys TGT	phe	pr CC1	369 cys TGC	4 59 59 59 59 59 59 59 59 59 59 59 59 59	asn 1 AAC (	\$60 gln leu CAG TTA	ser a TCA G	8 2 3
20		278 0ys 1GC	asp CAT	his CAT	동동				phe s TTT T	36 A
20	hls CAT	<b>1</b> 6	ala CCT	erg AGC	HIS CAT	aln leu CAG CTT	ser arg TCA AGA	leu e CTG A	cys p	6 13
	246 CY3 TGC	278 279 leu lys glu oys cys CTG AAG GAA TGC TGT	ela CCT	arg ACA	pro CCT	a se			pro cys phe CCA TGC TTT	518 11e cys thr leu ser glu lys glu arg ATA ICC ACA CTI ICT GAG AAG GAG AGA
	265 glu cys GAA TGC	leu CTG	leu TTA		asp CAT	phe TTT	glu val CAG GTC	val val Grc Grc	arg p CGA C	- 5 - 5
		ner ser lys TCC ACT AAA	ser ICA	tyr TAT		leu CTT		3er v 100 0	arg a	51& cys e TCC A
25	hla thr CAC ACG	Ber	leu pro TTG CCT	glu GAA	ala CCT	392 cys glu leu TGT GAG CTT	leu val CTT GTA	leu s CTA 1	asn arg AAC AGG	S 11e c ATA T
		100 P		tyr TAT	ala ecc	392 cys TCT	thr	tyr I	val a	asp 1 CAT A
	240 Iyo vol Aaa GTC	11e ATC	300 ala asp cct cac	leu 1TC	361 cys TGT		Pro	esp CAC 1	480 ser leu val TCC TTG GTG	ala asp CCA CAT
•		270 ser TCG	300 12 12 000 100 100 100 100 100 100 100 100 100	330 phe TTT	360 9ys 1GC	390 gin asn CAA AAT	\$20 thr ACT	450 91u 6 GAA 0	480 ser 1 TCC 1	510 hís a CAT G
	ch? Acc	gin asp CAA CAT	met pro ATG CCT	gly met GGC ATG	1ys AAG		ser TCA	ele ccA	glu s CAA 7	phe h TTC C
30	leu CTT		met ATG	91y CCC	glu thr thr leu glu GAA ACC ACT CTA GAG	ssn leu ile lys AAT TTA ATC AAA		448 0ys (	r g	
	thr asp ACA GAT	glu asn GAA AAT	g Se	leu TTG	leu CTA	leu TTA	pro gin val ccc cAA GTG	ຸຄຸ	thr lys cys cys thr ACC AAA TCC TGC ACA	phe t TTC A
•	ehr ACA	gle GAA	asp GAT	phe TTC	thr Act	asn	. ord	aet 1	476 477 cys cys TCC TCC	thr p
	۷۵ <b>۱</b> ۲۵	265 0ys TCT	asn AAT	val GTC	thr Acc	gla CAG		<b>5</b> 43	\$ %	2. A A A
35	lys leu AAG TTA	lys tyr lle AAG TAT ATC	val glu GTG GAA	asp CAT	2 g	pro gln CCT CAG	lys lys vel AAG AAA GTA	. ys 6	r SO	ala glu CCT CAA
	1ys AAG	tyr Tat	val GTG	lys asp AAG CAT	tyr TAT	gle Ga	Ys J	16 1 CA A	val t GTC A	asn a AAT G
	3e <b>r</b> 100		glu GAA	321 giu ala lys asp val phe leu GAG GCA AAG GAT GTC TTG	thr	ole GAA	thr 1	448 glu ala lys arg met pro cys GAA CCA AAA AGA ATG CCC TGT	arg val AGA GTC	phe ar
	231 val GTT	261 818 GCC	291 a1a GCC	321 91u GAG	351 1ys thr tyr AAG ACA TAT	381 val CTC	411 tyr thr lys lys vbl Tac acc aag aaa gta	Pro g	471 839 8 CAC A	501 glu phe asn ala glu thr phe thr GAG TTT AAT GCT GAA ACA TTC ACC

5	(1790)	ter TAA CATCACATTTAAAAG (1883)	lge lep Catetcagcetaccatgagaataagaaaatgaagatcaaaagcttattcatetgtttttegtttggtgtaaagccaacacectgtetaaaaacataaatttetttaa (2002)	
•	560 1ys AAG (	AG (	TAA (	
		ter TAAA	CTT)	
	558 559 0ys cys TGC TGC	ICATI	IATTI	
	lys o	ATCA	ATA	
10	glu lys	e G	IAAAC	
	al g	leu t TTA 1	TAA	
	ele ele phe vel GCT GCT TTT GTA	qly leu ter GCC TTA TAA	:TGTC	
	16 P	leu q TTA C	ACC	
9 F	18 8 CT C	ala leu GCC TTA	CAAC	
15	phe s	ila a	AAGC	
	550 asp phe CAT TTC	580 gln ala CAA GCT	;TGT/	
		er 9	7100	(078)
	met asp ATG CAT	Ta a	TTC	5
20	191 m	570 ala glu glu qly lys lys leu val ala ala ser GCC GAG GGT AAA AAA CTT GTT GCT GCA AGT	CTT	· · ·
	lya giu gin leu lys ala val AAÄ GAG CAA CTG AAA GCT GTT	al e		: 2
	8. X	. ne .	וכדפו	•
	eu 1 :TG A	84 6 8	ITCAI	\$
	AA C	ys N	TTA	Arct/
25	34 G	5 5	AAQ	AGA
	<b>\$ \$</b>	ale o	VTCA/	892
	540 thr 1 ACA A	570 glu g GAG G	*AAG	MAA!
	ala t GCA A	2 E S	WATC	T.
30	lys ala AAG GCA	phe a	¥6€	ATTA
30		567 573 150	88	TCA
	lys pro AAG CCC	4 5	ter Itaa	3760
	16 J	) ole	ter ter Itcacaataa	STCT
	AA C	lys (	icate	ITTE
35	leu val lys his CCC CTC GAG AAA CAC	567 asp asp lys glu thr cys GAC GAT AAG GAG ACC TGC	CTAC	TCATTTTGCCTCTTTTCTCTGTGCTTCAATTAAAAAATGGAAGAATCTAA 20AA (2078)
	יונה ע זרה ה	ssp a	rcace	71100
	531 glv 1 GAG C	561 818 6 CCT (	:ATC1	ICATI
	50			-

9. Nucleotide sequence coding for the pre pro human serum albumin, said nucleotide sequence is as follows:

5	(30)	(170)	(560)	(350)	(088)	(330)	(620)	(710)	(300)
	Ber Age	20 178 AAA	50 ala GCA	Be CTT	110 Pro CCA	140 Cry TAT	170 qln CAA	200 cys TGT	23 410 6AA
	leu ohe CTC TTT	phe TTC	alu phe CAA TTT	ala thr GCA ACT	asn	leu TTA	169 cys 1CC	1ys AAG	ele GCA
		AAT			ase CAC	tyr TAC	168 cys TGT	leu CTC	phe TTT
10	-10 lle ser leu leu phe ATT TCC CTT CTT TTT	alu alu GAA GAA	thr ACT	75 leu oys thr val TTA TCC ACA GTT	ohe leu qin his iys əsd TTC TTG CAA CAC AAA GAT	qlu thr phe leu lys lys CAG ACA TTT TTG AAA AAA	alu GAA	ala iya ain ara GCC AAA CAG AGA	atu GAG
	lau CTT		glu val GAA GTA	thr ACA	lys Aaa	1ys AAA	th. Aca	al cac	ele C:T
	lau CTT	leu aly TTG GGA	SA A	2 2 3 3	gin his Caa cac	leu TTG	phe TTT	Lys AAA	phe pro lys ala TTT CCC AAA GCT
	38E		asn AAT	leu TTA	<b>ala</b> CAA	phe	ele CCT	318 CCC	5 J
	Bleatt	IYO OSD AAA CAT	leu val TTA GTG	asp lys CAC AAA	leu 17G	th. ACA	ala GCT	ser ICT	phe TTT
15	pho TIT	8yo AAA		28 28 28	ohe TTC	alu GAG	eyf lys tat aaa	ser 7CG	arg ACA
	e e o tep vol tlu tgc gta acc	phe TTT	80 Vel lys GTA AAA	70 phe gly asp lys TTT CGA CAC AAA	100 101 alu oys CAA TCC	e as	tyr TAT	190 lys ala AAC CCT	220 ser gin arg ACC CAG AGA
	e vol	<b>်</b> စု <sup>က္</sup> အသ	%0 Vel GTA	70 phe 111			160 Pre ACC	130 173 AAG	220 ser ACC
	2 2 2 2 2	his Cat	h is CAT	leu	esn AAT	asa GAC	ala lys GCT AAA	alu aly CAA CCC	leu CTG
20	8 yo AAG	olo CCT	esp CAT	thr ACC	<b>P</b> 8	his CAT	ala CCT	<b>2</b> 8	ဦး
20	-18 Kat ATG	glu vol CAC CTT	g S	leu hís CTT CAT	919	ala phe CCT TTT	phe ohe TTC TTT	asp GAT	ala GCT
			phe TTT		SCT TO			<b>5</b> 20	vel GTA
		lyo der aag agt	<b>2</b> 23	ser ICA	gle GAA	thr	leu CTT	leu CTT	trp ala TGG GCA
			3& cys 1Gf	1ys AAA	c AA	12& 073 1GC	leu CTC	glu GAA	£ 55
25		olo his GCA CAC	gla	dse CAC	Nys AAA	aret AT <b>G</b>	glu	asp GAT	lys ele AAA GCA
		o) စ <b>ငင်</b> A	gla	62 cys TGT	515 GCA	val	970 200	leu CTC	1ys Aaa
		oop Cat	leu CTT	AAT	90 91 cys cys ICC ICI	asp GAT	ala တင	180 pro lys cca Aag	210 ala phe GCT TTC
		2- 8-0 8-0 8-0 8-0 8-0 8-0 8-0 8-0 8-0 8-	30 tyr TAT	8 28		120 val GTT	150 tyr TAT	180 Pro CCA	210 ala phe GCT TTC
		org	gh	ala GCT	asp GAC	g]u GAG	phe TTT	leu leu CTG TTG	P S
30		ghe TTT	ola GCT	ser 1CA	ala R	<b>5</b> 50	tyr TAC	leu CTG	g Per
		gly vol cet cte	phe TTT	9]v 686	glu met GAA ATG	arg ACA	pro CCT	177 eys 100	91. 88.
		gly cct	ala OCC	ese CA7	g) GAA	val GTG	h1s CAT	ele cc	phe TTT
		4 2 3 A 3 A 3 A 3 A 3 A 3 A 3 A 3 A 3 A 3	11e ATT	ele CCT	91y ccr	leu TTG	arg AGA	ala CCT	1ys
35		 Ser 70	leu TTG	val	tyr Tat	pro arg leu val arg CCC CGA TTG GTG AGA	ala arg arg his pro CCC AGA AGA CAT CCT	1ys AAA	를 A
		olo tyr oer CCT TAT ICC	leu val TTG GTG	53 thr cys ACA TGT	chr ACC	) ) )	ala OCC	ala esp lys ele ele oys GCT GAT AAA GCT GCC TGC	201 ala ser leu gln lys phe gly GCC AGT CTC CAA AAA TTT GCA
		۵ <b>۱</b> ۵ <b>۵۲</b> ۲	Je 116	S ACA	2 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8	leu CTC	1116	ala GCT	ser
		00P	21° 21° 30°	51 173 AAA	91 879 CGT	111 asn AAC	142 CAA /	171 ala CCT	201 818 CCC

5	(890)	(980)	(1070)	(1160)	(1250)	(1340)	(1430)	(0251)	(1610)	530 val GTT (1700)
	260 leu CTT	290 11e Att	320 ele CC:T	350 818 GCC	380 1ev CCT	&10 ara CGT	840 N13 CAT	470 ser AGT	5n0 1ys AAA	530 val GTT
	889 CAC	289 698 100	tyr Tat	leu	pro CCT	val CTT	438 cys lys TCT AAA	pro val CCA GTA	57 222	leu CTT
	esp arg ele d	h is CAC	AAC	leu ara CTG AGA	phe lys ITT AAA	ala leu leu val GCG CTG TTA GTT	438 cys TCT	<b>52</b> 0	val CTT	e) e GCA
10	era AGG	lys ser AAA TCC	lys AAA	leu CTG	phe III	leu CTG	437 cys TCT	the Acc	alu thr tyr val GAA ACA TAC GTT	thr
	685 GAC		316 cys 700	leu leu CTG CTG	GAA		1 ys	elu lys	thr ACA	E &
	ola esp CCT CAT	leu glu TTC CAA	vel	ser val val leu leu aeu leu TCT GTC GTG CTG CTG AGA CTT	phe asp (	gln esn CAG AAT	437 438 qly ser lys cys cys lys GCC AGC AAA TGT TGT AAA	460 461 leu oys val leu his glu lys TTA TGT GTG TTG CAT GAG AAA	olu GAA	520 glu arq ain ile lys lys ain thr als leu val GAG AGA CAA AIC AAG AAA CAA ACT GCA CTT GTT
			CAT	val GTG	phe TTC	e la CAG		leu his TTG CAT	asp	)ys AAG
	293 głu cys Gaa ycy	pro leu CCT CTG	Ber lys AGT AAG	val GTC	lys val AAA GTG	phe TTC	val G7G	leu TTG	glu val asp GAA GTC GAT	11e ATC
15			Be <b>r</b> A <b>GT</b>		1 ys	Bys AAA	430 leu qly lys val CTA GGA AAA GTG	val	9 4 6 8 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6	E S
	250 leu leu CTG CTT	280 qlu lys Gaa aaa	g g GAA	tyr TAC	ရ လိုင်	400 glu tyr GAG TAC	4) b	461 999 161	890 ala leu CCT CTG	AGA AGA
			310 vel CTT	340 889 GAT	370 tyr TAT		430 3eu CTA			520 91u GAG
	gly sop cca cat	278 279 cys cys TCC TCT	esp phe CAT TTT	his pro CAT CCT	369 qlu cys CAA TGC	<del>2</del> 8	erg esn ACA AAC	leu asn gln CTG AAC CAG	ser TCA	glu lys CAC AAG
20	914 664	278 278 160		h1s CAT	glu GAA	leu CTT	ACA	asn AAC	phe TTT	
20	M80 CAT	e SA	919	arg AGG	A TA	alu aln GAG CAG	ser 1CA	Jeu CTG	pro eys CCA TGC	514 oys thr leu ser TCC ACA CTT TCT
	% % % % % % %	lys AAG	ala GCT	ala arg GCA AGA	8 155	a Ju GAG	glu val GAG GTC	val GTC		leu CTT
	243 1970	leu CTG	leu TA	ele GCA	SP	phe TTT		val GFG	දී වි	thr
	glu GAA	1ys	ser ICA	glu tyr GAA TAT	978 808	leu CTT	thr leu val ACT CTT GFA	ser 700	arg AGG	514 0y8 100
25	hio the Cac acg	ser	leu pro TTG CCT		als CCT	alu GAG	leu CTT	leu CTA	asn AAC	11e ATA
		ser TCC		tyr TAT	#1# CCC	392 cys TGT	thr	tyr Tat	val GTG	GAT
	280 Iyo voi Aaa GTC	11e	6 8 5 C 8 5	leu 11G	361 cys TGT	asn	S A	450 glu asp GAA GAC	480 ser leu TCC TTG	510 his ale cat cca
		270 ser TCG	300	330 phe TTT	360 978 750	3% gln CAA	\$20 thr ACT			
	EP. ACC	GAT	pro CCT	gly met GCC ATG	glu lys GAG AAG	<b>1</b> 58	ser TCA	ala ccA	glu GA	phe TTC
30	lew CTT	cA CA	met ATG			11e ATC	val GTG	448 cys 1GT	thr ACA	thr
	30g	asn AAT	914	phe leu TTC TTG	leu CTA	leu TTA	gla CAA	Ret pro ATG CCC	477 0ys 1GC	phe TTC
	S ACA	glu	glu asn asp GAA AAT GAT	phe TTC	glu thr thr leu GAA ACC ACT CTA	pro gin asn leu CCT CAG AAT TTA		met ATG	476 0ys 1CC	thr ACA
	vol.	265 11e cys ATC TGT	glu asn GAA AAT	val	thr	g CAG	vol	arg AGA	val thr lys GTC ACC AAA	asn ala glu AAT GCT GAA
35	leu TTA			asp GAT	glu AA	pro	1ys AAA	1ys AAA	thr	al a GCT
	ser byo leu vol thr oop TCC AAG TTA GTG ACA GAT	tyr TAT	glu val GAA GTG	ala lys asp val GCA AAG GAT GTC	tyr TAT	glu GAG	thr lys lys vol pro gln ACC AAG AAA GTA CCC CAA	glu ele lys erg met pro GAA GCA AAA AGA ATG CCC	val GTC	phe asn ala glu thr phe IIT AAI GCT GAA ACA ITC
		1ys		ala GCA	ACA ACA	glu GAA	thr ACC	ale GAA	4 P P P	phe TTT
	239 Vol CTT	261 ala GCC	291 818 GCC	321 glu GAG	351 1ys AAG	381 val GTG	411 tyr TAC	441 pro CCT	471 asp GAC	501 914 GAG

	(1790)	leu ter TFA TAA CATCACATTTAAAAG (1883)	ter tep Catcicacciaccatachgaagaaaatgaagattattititititititititititi
	S60 Nys	AG .	
	559 Cys 100	ter ITAA	
	558 678 700	ACAT	
	1 ys	CATC	6
	glu GAG	ter TAA (	4
	val Gta	leu TTA	
	phe TTT	<u> </u>	300
	ela GCT	leu TTA	Č
•	ala ect	္ ၁၂၂	او
	phe 71C	ela Scr	7
	558 559 Iyo giu gin leu lys ala val met asp asp ahe ala ala phe val giu lys cys cys AAA GAG CAA CTG AAA GCT GTT ATG GAT GAT TTC GCT GCT TTT GTA GAG AAG TGC TGC	570 phe sla glu glu qly lys lys leu vol olo sla ser glm sla slo leu qly leu ter TTT GCC GAG GGT AAA AAA CTT GTT GCT GCA AGT CAA GCT GCC TTA GGC TTA TAA	
	esp CAT	ser AGT	0440
	net ATG	919 6CA	. 444
	val GTT	ol o CCT	77.
	ala GCT	vəl	8888
•	1 ys	leu CTT	7.01
	lau CTG	Iys AAA	TTCA
	gln CAA	1 ys AAA	27.7.0
	<b>9</b>	4 y 667	70000
	1 yo	<b>9</b> 60	7380
	<b>540</b> thr	szo glu Gac	CAAC
	ela GCA	ချော	MAAA
	1ys AAG	phe TT	AAG.
	<b>22</b>	567 593 160	586
	lys AAG	the	tes WTA
_	his lys cac aag	age Geo	ter ter Tcacaataa
•	1ys	1ys AAG	CCA1
	leu val lys P CTC GTG AAA C	esp GAT	CCTA
	leu CTC	asp asp GAC GAT	TCAC
	531 91u CAC	561 212 CCT	27.5

ICATITIGCCICTITICICIGCCTICAATIAATAAAAATGGAAAGAATCTAA..... 20 .....AA (2078)

- 10. A nucleotide sequence according to any of claims 6 to 9, in essentially pure form.
- 11. A DNA transfer vector comprising a nucleotide sequence as defined in claim 5.
- 5 12. A DNA transfer vector according to claim 11, transferred to and replicated in a micro-organism.
  - 13. A DNA transfer vector according to claim 12, which is a plasmid.
- 14. A DNA transfer vector according to claim 13,10 wherein the plasmid is pBR322 or YEp6.
  - 15. A process for preparing human serum albumin, which comprises culturing a micro-organism according to claim 5.
  - 16. A DNA transfer vector according to any of
- 15 claims 12 to 14, or a process according to claim 15, wherein the micro-organism is a bacterium or yeast.
  - 17. A vector or process according to claim 16, wherein the bacterium or yeast is <u>E. coli</u> or <u>Saccharomyces cerevisiae</u>.

1/1 Mpo (1) 73613) 37.√ 3.√ © % Z Z ۵. 586 ferm TAA Restriction Endonuclease Map of Human Serum Albumin cDNA Clones 493/4 Tog ! 419/0 450/1 Ninc II Mbo II ۵. ه <u>ک</u> ھ pMA36 382 Sibo 8 ک الا Kilobasos 325/6 Mbe II ဓာ့ Pat i (3611) Hpa II (3548) 182/3 Tog 1 ณ์ Ø, Hpa II (3548) pHA206 က 57 Hini I Pst 1 Mbo II Mbo II (3611) 16/7 31 જ Hpa II (3658)